

A Minimal Model for Predicting Visual Search in Human-Computer Interaction

Tim Halverson and Anthony J. Hornof

Department of Computer and Information Science
1202 University of Oregon, Eugene, OR 97403-1202 USA
{thalvers, hornof}@cs.uoregon.edu

ABSTRACT

Visual search is an important part of human-computer interaction. It is critical that we build theory about how people visually search displays in order to better support the users' visual capabilities and limitations in everyday tasks. One way of building such theory is through computational cognitive modeling. The ultimate promise for cognitive modeling in HCI is to provide the science base needed for predictive interface analysis tools. This paper discusses computational cognitive modeling of the perceptual, strategic, and oculomotor processes people used in a visual search task. This work refines and rounds out previously reported cognitive modeling and eye tracking analysis. A revised "minimal model" of visual search is presented that explains a variety of eye movement data better than the original model. The revised model uses a parsimonious strategy that is not tied to a particular visual structure or feature beyond the location of objects. Three characteristics of the minimal strategy are discussed in detail.

Author Keywords

Visual search, cognitive modeling, cognitive strategies, eye movements, screen design, EPIC.

ACM Classification Keywords

H.5.2. [Information Interfaces and Presentation]: User Interfaces – Evaluation/methodology, graphical user interfaces (GUI), screen design, theory and methods; H.1.2 [Models and Principles]: User/Machine Systems; I.2.0 [Artificial Intelligence]: General – Cognitive simulation.

INTRODUCTION

Visual search is an important part of human-computer interaction (HCI). For most users and many tasks, information is obtained through visual search. The visual search processes people use in these tasks have a substantial

effect on the time to find and the likelihood of finding the information they seek.

One way to better understand the visual search processes people use, and why they use them, is with computational cognitive modeling. Theory developed through cognitive modeling, as is done in this research, is essential for the development of automated interface analysis tools. Interface designers can use such tools to evaluate visual layouts early in the design cycle before user testing. Two tools that could benefit from a straightforward, minimal model of visual search are CogTool [6] and G2A [11].

There are many cognitive models of visual search that may one day converge to form a solid basis for the theory of visual search in HCI [1,4,12]. While these models are very useful, many such models are designed to explain the effects of particular visual structures or salient features. The research reported here is motivated by the need to find a minimal model for goal-directed visual search that is not tied to a particular visual structure or feature saliency. A minimal model of visual search is presented that explains a variety of eye movement better than previous research of the same task.

PREVIOUS RESEARCH

This work builds on previous modeling and eye movement analysis of menu search. Hornof [4] studied the visual search of layouts with and without a visual hierarchy and built computational cognitive models of the task. Hornof and Halverson [5] replicated the study to evaluate the eye movement strategies predicted by the models and found that while the models predicted the search time and a fair amount of the visual search behavior, some critical aspects of the visual search behavior (for example, scanpaths) were not well predicted. A goal of the current research is to improve the original models by accounting for more eye movement data found in the follow-up study.

Figure 1 shows the task relevant to the current research (the "unlabeled" layouts from [4,5]). Sixteen participants searched four different screen layouts for a precued target. Each layout contained one, two, four, or six groups. Each group contained five objects. The groups always appeared at the same locations on the screen. One-group layouts used group A. Two-group layouts used groups A and B. Four-group layouts used groups A through D.

To appear in the Proceedings of the Conference on Human Factors in Computing Systems, San Jose, CA, April 28-May 3, 2007.

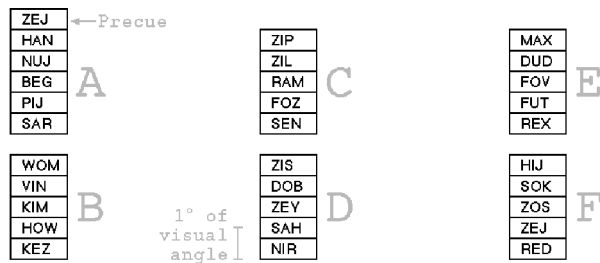


Figure 1. A 6-group layout. The precue, in the top left, would disappear when the layout appeared. The gray text did not appear during the experiment.

In the original models, the simulated eyes moved down the first column of text, then down the second column, and then down the third. Furthermore, the eyes jumped over a carefully controlled number of items with each eye movement. The model accounted for the reaction time and a fair number of eye movement measures, considering that the model was built without eye movement data to guide its development.

However, the model's strategy is somewhat tuned to aspects of this one visual task and layout. The model directly controls the direction and amplitude of eye movements. This direct control, while providing a good fit to the reaction time data, does an unsatisfactory job of explaining people's visual scanpaths. The original model did a better job of predicting the frequency and number of fixations, but there is room for substantial improvement. A goal of this research is to improve the accuracy with which the model explains people's visual search strategies, while at the same time maintaining a minimal model that does not directly control the scanpaths based on the visual structure of the layout or visual properties of the layout items.

CHARACTERISTICS OF A MINIMAL MODEL

This research proposes three characteristics of a minimal model of visual search: (a) Eye movements tend to go to nearby objects, (b) fixated objects are not always identified, and (c) eye movements start after the fixated objects are identified. These characteristics are motivated by previous research and eye movement data, and are introduced to the model here in a step by step manner. We propose that any applied model of visual search should include at least these three characteristics, and furthermore that much visual search behavior can be explained by the integration and interaction of these three characteristics.

The cognitive models described in this study were built using the EPIC (Executive Process-Interactive Control) cognitive architecture [7]. EPIC captures human perceptual, cognitive, and motor processing constraints in a computational framework that is used to build cognitive models. EPIC simulates ocular-motor processing, including the ballistic eye movements known as saccades and the fixations during which the eyes are stationary and information is perceived. Visual properties of objects are available at varying eccentricities and timing. For example,

the text property is available within one degree of visual angle from center of fixation and arrives in working memory 150ms after the text is fixated.

The minimal model was derived iteratively by making gradual improvements to the model based on eye movement data. At each step in the model's development, a sub-strategy was added or a perceptual parameter was changed to increase the model's fidelity.

A potential criticism of the task modeled here is that it lacks ecological validity and any change to the task may invalidate the resulting model. We acknowledge this concern but point out that the model captures fundamental human perceptual-motor processes, capabilities and constraints that will be common across a wide range of ecologically valid, real-world tasks such as air-traffic control. Common processes and constraints include error in object recognition, biases towards shorter saccades, and fixation duration control.

The resulting model is useful for predicting visual search in HCI. The model contains a visual search strategy that is not tied to a particular visual structure or saliency of a feature beyond the location of the visual objects. A text feature is used to determine if the target is found, but does not guide search. The development of the model and the integration of the three key characteristics are discussed next.

Eye Movements Tend to Go to Nearby Objects

The basic job of the human visual search process is to decide which objects to fixate. Though a completely random search strategy is very useful for predicting the mean layout search time, people do not search completely randomly. Instead, people enjoy the many benefits of moving to objects that are relatively nearby rather than across the layout. Saccade destinations tend to be based on proximity to the center of fixation [9].

In the current research, rather than searching randomly or following a prescribed search order (as in the original model), a strategy was used that selects saccade destinations with the least eccentricity. To account for variability in the human saccade distances, noise is added to the model's process of selecting the next saccade destination as follows: (a) After each saccade, the *eccentricity* property (the distance from the eye position) of all objects is updated based on the new eye position. (b) The eccentricity is scaled by a fluctuation factor, which has a mean of one and a standard deviation of 0.3. This scaling factor is individually sampled for each object. (c) Objects whose text has not been identified and that are in unvisited groups are marked as potential saccade destinations (i.e. search without replacement). (d) The candidate object with the lowest eccentricity is selected as the next saccade destination.

The standard deviation of the fluctuation factor was determined by varying the fluctuation factor to find the best fit of the mean saccade distance. As shown in Figure 2, the

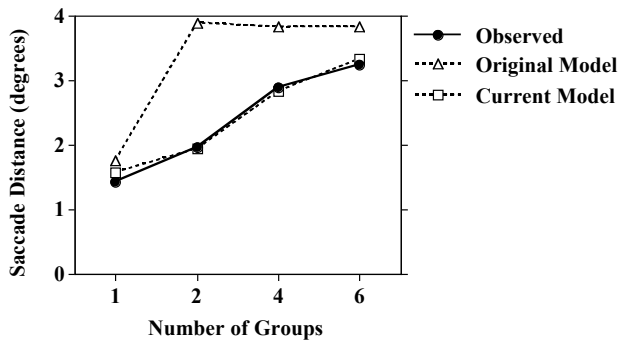


Figure 2. Saccade distance observed (circles), predicted by the original model (triangles), and predicted by the current model (squares). The standard errors of the observed data are too small to be visible.

Original model AAE = 43.3%. Current model AAE = 4.2%.

current model predicts the mean saccade distances very well, with an average absolute error (AAE) of 4.2%, a considerable improvement over the AAE of 43.3% in the original model. This strategy also does a good job of predicting the observed scanpaths. Figure 3 shows the three most frequently observed scanpaths, and how the current model predicts the observed scanpath frequencies better than does the original model.

This “nearby with noise” strategy used in the minimal model has a couple of benefits for predicting visual search compared to models tied to particular visual structures or saliency of visual features. First, only the location of the layout objects if required. This is beneficial if other properties in the layout are unknown or difficult to extract. Second, this search strategy can be used when the visual saliency alone cannot predict visual search, as is the case with goal-directed search [8]. Unlike the original model [4], this minimal model does not require a predefined notion of how the eyes will move through the layout to predict the observed scanpaths.

Fixated Objects are Not Always Identified

One goal of the current research was to produce a model that accounts for multiple eye movement measures. Although a model that moves the eyes to nearby items accounts for the observed scanpaths, improvements were required that accounted for the observed number of fixations per trial.

Studying the eye movement data, it was found that participants sometimes fixated on or near the target but

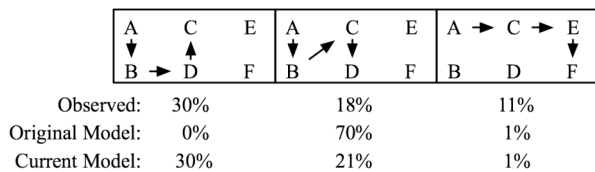


Figure 3. The most commonly observed scanpaths in six-group layouts and how often each path was taken by the participants (observed) and the models (original and current).

continued to search. This suggests that the participants may occasionally fail to recognize the target, even though they eventually complete the trial correctly.

Previous modeling research [2,10] suggests that people do occasionally fail to recognize fixated text. The minimal model was modified to include a *text recoding failure rate*. Text recoding failure rate is a recent addition to EPIC, and the default value is zero (i.e. no chance of failing to identify text). The parameter represents the probability that the text property of an object will not be encoded.

The text recoding failure rate parameter was used in the current work for two reasons. First, to explore ways to account for the observation that participants missed the target occasionally. Second, if the current modeling predicts observed eye movement data with a text recoding failure rate similar to that used in the previous modeling, this would not only support the use of the parameter here, but also suggest a recommended default value for the parameter in future modeling efforts.

The text recoding failure rate was initially set to 10%, the value used in [2]. This failure rate was changed by 1% increments until the model predicted the mean number of fixations per trial. A value of 9% provided the best fit for the number of fixations per trial.

As shown in Figure 4, the current model predicts the number of fixations per trial very well, with an AAE of 4.2%. This is an improvement over the original model [4] and the current model with no text recoding failure rate. The decreased error and the similarity between the best-fitting text recoding failure rate found here and the rate found in past research provides support for the use of the text recoding failure rate parameter. Again, we are maintaining a minimal model in that this improvement to the model does not require layout-specific information. Future research will need to address the possibility of encoding failure rates for non-text stimuli.

Eye Movements Start After Objects are Identified

The underlying concepts of the minimal search strategy

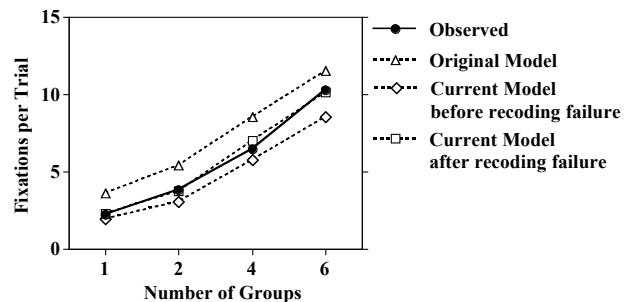


Figure 4. Fixations per trial observed (circles), predicted by the original model (triangles), predicted by the current model with 0% encoding failure (diamonds), and predicted by the current model with 9% encoding failure (squares). Original model AAE = 37.8%. Current model with 0% encoding failure AAE = 14.3%. Current model AAE = 4.2%.

developed thus far have specified *what* the eyes move to. A visual search strategy also needs to specify *when* the eyes move. Various strategies have been proposed for how long people fixate items (see [3] for an overview). The two basic competing theories are (a) *preprogramming*, in which fixation durations are directly controlled by the search strategy, and (b) *process-monitoring*, in which fixation durations are determined by the time required to perceive the fixated stimuli. The minimal model utilizes a process-monitoring strategy, which requires fewer production rules and parameters than required by a preprogramming strategy.

In the model, saccades are initiated after objects in the fovea are identified. Once the simulated eyes reach their destination, the strategy waits until the *text* property of the fixated objects is available. While waiting, the strategy starts the process of deciding where the eyes will go next.

As shown in Figure 5, the current model predicts the fixation durations very well, with an AAE of 4.6%. This is an improvement over the original model [4] that had an AAE of 26.5%. The use of a process-monitoring model for determining fixation durations predicts the observed data very well.

CONCLUSION

The minimal visual search model discussed here will be useful to further research in predicting and understanding user behavior in HCI. Such a model could be used in future cognitive modeling as a base on which to build more robust models of visual search. Further, predictive tools like CogTool [6] could incorporate a similar model for predicting users' visual search behavior. Theory developed through cognitive modeling such as the work presented here is essential for the development of predictive, automated interface analysis tools that allow designers to evaluate their visual layouts early in the design cycle before user testing is feasible.

A minimal model of visual search accounts for a variety of eye movement data, from fixation duration to the most common scanpaths. The model does so primarily by

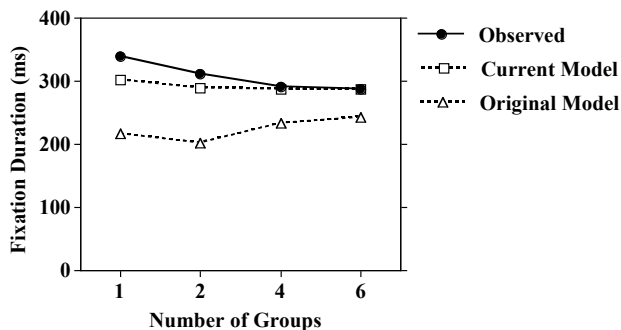


Figure 5. Fixation duration observed (circles), predicted by the original model (triangles), and predicted by the current model (squares).

Original model AAE = 26.5%. Current Model AAE = 4.6%

employing three straightforward characteristics, motivated by eye movement data and previous research, that can be applied to other modeling research. These principles are: (a) Eye movements tend to go to nearby objects, (c) fixated objects are not always identified, and (d) eye movements start after the fixated objects are identified. This minimal model does a better job of accounting for the observed visual search behavior than a previous model of the same task that was not informed by eye movement data.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research and the National Science Foundation.

REFERENCES

1. Brumby, D. P., & Howes, A. (in press). Strategies for Guiding Interactive Search: An Empirical Investigation into the Consequences of Label Relevance for Assessment and Selection. *Human-Computer Interaction*.
2. Halverson, T., & Hornof, A. J. (2004). *Explaining Eye Movements in the Visual Search of Varying Density Layouts*. In Proc. ICCM 2004, Lawrence Erlbaum Associates, 124-129.
3. Hooge, I. T. C., & Erkelens, C. J. (1996). Control of fixation duration in a simple search task. *Perception and Psychophysics*, 58, 969-976.
4. Hornof, A. J. (2004). Cognitive Strategies for the Visual Search of Hierarchical Computer Displays. *Human-Computer Interaction*, 19(3), 183-223.
5. Hornof, A. J., & Halverson, T. (2003). *Cognitive strategies and eye movements for searching hierarchical computer displays*. In Proc. CHI 2003, ACM Press, 249-256.
6. John, B. E., Prevas, K., Salvucci, D. D., & Koedinger, K. (2004). *Predictive Human Performance Modeling Made Easy*. In Proc. CHI 2004, ACM, 455-462.
7. Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12(4), 391-438.
8. Koostra, G., Nederveen, A., & de Boer, B. (2006). *On the Bottom-Up and Top-Down Influences of Eye Movements*. In Proc. Proc. CogSci 2006, Lawrence Erlbaum Associates, 2538.
9. Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research*, 38(12), 1905-1815.
10. Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science*, 12(4), 287-292.
11. St. Amant, R., & Ritter, F. E. (2004). *Automated GOMS-to-ACT-R Model Generation*. In Proc. ICCM 2004, Lawrence Earlbaum, 26-31.
12. Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1(2), 202-238.